

Komparasi Algoritma *Machine Learning* untuk Klasifikasi Kunjungan Tamu pada Institusi Pendidikan

^{1*}Andi Laksono, ²Rony Heri Irawan, ³Umi Mahdiyah

^{1,2,3}Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ¹andilaksana125@gmail.com, ²ronyag305ku@gmail.com, ³umimahdiyah@gmail.com

Penulis Korespondens : Andi Laksono

Abstrak—Pencatatan tamu di institusi pendidikan umumnya sekadar arsip pasif, belum dimanfaatkan untuk prediksi kategori kunjungan. Riset komparasi *machine learning* juga masih minim di domain operasional ini. Studi ini membandingkan algoritma KNN, Naive Bayes, Decision Tree, Random Forest, dan SVM dalam mengklasifikasikan data tamu menggunakan Orange Data Mining. Pemilihan platform visual ini memudahkan penerapan oleh staf non-programmer. Hasil penelitian menunjukkan algoritma paling optimal untuk kasus tersebut. Secara praktis, studi ini membuktikan bahwa machine learning mampu mengubah data administratif pasif menjadi wawasan prediktif strategis bagi pengambilan keputusan dan peningkatan layanan institusi pendidikan.

Kata Kunci—Klasifikasi, Kunjungan Tamu, Orange Data Mining, Pembelajaran Mesin, Perbandingan Algoritma

Abstract—Guest logging in educational institutions is generally limited to passive archiving and has not been utilized for predicting visit categories. Comparative machine learning research also remains limited in this operational domain. This study compares KNN, Naive Bayes, Decision Tree, Random Forest, and SVM algorithms in classifying guest data using Orange Data Mining. The selection of this visual platform facilitates implementation by non-programmer staff. The results indicate the most optimal algorithm for this specific case. Practically, this study demonstrates that machine learning can transform passive administrative data into strategic predictive insights, supporting decision-making and enhancing services in educational institutions.

Keywords—Algorithm Comparison, Classification, Guest Visit, Machine Learning, Orange Data Mining

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Institusi pendidikan merupakan lingkungan yang dinamis yang tidak hanya berinteraksi dengan sivitas akademika internal, tetapi juga dengan berbagai pihak eksternal. Setiap harinya, terjadi arus kunjungan tamu yang masuk ke lingkungan kampus atau sekolah, baik itu wali murid, tamu dari instansi pemerintah, perusahaan mitra, vendor, alumni dan masyarakat umum. Pengelolaan data kunjungan tamu ini merupakan aspek penting dalam tata kelola administrasi dan

keamanan institusi. Namun, pada praktiknya, pencatatan kunjungan tamu di banyak institusi pendidikan masih bersifat konvensional dan hanya berfungsi sebagai arsip pasif. Data yang terkumpul belum dimanfaatkan secara optimal untuk mengekstrak informasi atau pola yang berharga, seperti memprediksi kategori tujuan kunjungan atau mengklasifikasikan profil tamu secara otomatis. Padahal, informasi ini sangat krusial untuk membantu pihak keamanan dan administrasi dalam mengambil keputusan secara cepat, mengalokasikan sumber daya, serta meningkatkan kualitas layanan.

Seiring dengan perkembangan teknologi, pendekatan berbasis data menggunakan teknik *Machine Learning (ML)* telah menjadi solusi efektif dalam mengolah data besar menjadi pengetahuan yang bermakna (*knowledge*). Salah satu teknik ML yang paling sering digunakan untuk memprediksi dan mengkategorikan data adalah klasifikasi. Berbagai algoritma klasifikasi telah terbukti mampu memberikan akurasi yang tinggi di berbagai domain. Di bidang kesehatan, algoritma klasifikasi berhasil memprediksi penyakit kanker payudara [1], penyakit jantung [2], status stunting [3], tingkat obesitas [4], hingga penyakit stroke [5]. Sementara itu, di bidang akademik, algoritma ini juga diaplikasikan untuk memprediksi kelulusan mahasiswa [6][7], tingkat depresi mahasiswa [8], hingga prediksi calon mahasiswa baru yang mundur [9].

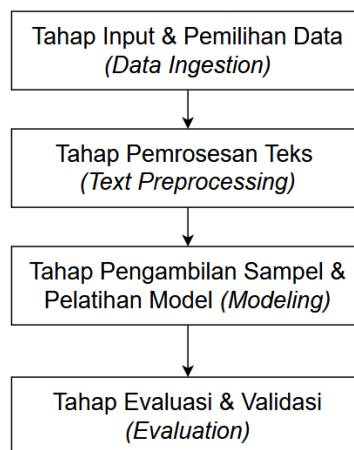
Meskipun demikian, tidak ada satu algoritma klasifikasi yang mutlak paling unggul untuk segala jenis dataset. Kinerja algoritma sangat bergantung pada karakteristik data, jumlah fitur, dan distribusi kelas pada dataset tersebut. Penelitian-penelitian sebelumnya menunjukkan hasil komparasi yang beragam. Sebagai contoh, penelitian terkait klasifikasi penyakit jantung dan obesitas menunjukkan bahwa algoritma *Random Forest* memiliki kinerja terbaik dibandingkan algoritma lainnya [2][4]. Di sisi lain, pada prediksi kelulusan mahasiswa, *algoritma K-Nearest Neighbor (KNN)* justru mengungguli algoritma lainnya dengan akurasi tertinggi [6][7]. Bahkan, pada domain prediksi stunting, *Naive Bayes* terbukti paling optimal [3]. Hal ini membuktikan bahwa pemilihan algoritma yang tepat memerlukan proses komparasi [10].

Berdasarkan kajian literatur tersebut, terlihat adanya gap penelitian (*research gap*). Mayoritas penelitian komparasi algoritma klasifikasi saat ini masih terkonsentrasi pada domain kesehatan dan akademik (performa mahasiswa). Riset yang mengaplikasikan teknik Machine Learning secara spesifik pada data operasional institusi pendidikan, khususnya untuk klasifikasi kunjungan tamu, masih sangat jarang ditemui. Padahal, data kunjungan tamu memiliki karakteristik unik yang berpotensi menghasilkan pola tertentu yang bisa diklasifikasikan, misalnya berdasarkan tujuan kunjungan atau kategori tamu yang melakukan kunjungan.

Oleh karena itu, untuk mengisi celah penelitian tersebut, studi ini bertujuan untuk melakukan komparasi kinerja beberapa algoritma Machine Learning populer seperti *K-Nearest Neighbor (KNN)*, *Naive Bayes*, *Decision Tree*, *Random Forest* dan *SVM*, dalam mengklasifikasikan data kunjungan tamu pada institusi pendidikan. Proses pemodelan dan evaluasi pada penelitian ini akan menggunakan pendekatan visual berbasis Visual Programming melalui aplikasi *Orange Data Mining* [11]. Penggunaan tools ini dipilih karena kemampuannya dalam menyediakan alur kerja (*workflow*) yang interaktif dan visual, sehingga proses Data Mining dari tahap preprocessing, pemodelan, hingga evaluasi dapat direplikasi dengan mudah oleh pihak administrasi institusi yang mungkin tidak memiliki latar belakang pemrograman yang mendalam [6].

Hasil dari penelitian ini diharapkan tidak hanya dapat memberikan rekomendasi model algoritma terbaik untuk kasus klasifikasi kunjungan tamu, tetapi juga menjadi bukti empiris baru (novelty) bahwa teknik komparasi Machine Learning dapat diaplikasikan secara efektif pada domain operasional administrasi institusi pendidikan untuk mengubah data pasif menjadi informasi yang bernilai strategis.

II. METODE



Gambar 1. Alur Kerja Penelitian

Workflow Orange Data Mining pada gambar di atas merupakan alur kerja untuk *Text Mining* (Analisis Teks) dan Klasifikasi (*Machine Learning*). Secara umum, alur ini mendemonstrasikan proses membaca data teks, melakukan pembersihan teks, mengubah teks menjadi fitur numerik, menguji beberapa model klasifikasi, dan mengevaluasi kinerjanya. Berikut adalah penjelasan detail untuk setiap tahapan alur (*workflow*) tersebut, dibagi menjadi 4 bagian utama:

2.1. *Data Ingestion* (Akuisisi & Seleksi Data)

Tahap awal penelitian berfokus pada pengumpulan dan penyiapan data mentah. Dokumen teks diimpor ke dalam sistem, disaring melalui visualisasi data awal, dan dilakukan pemisahan peran kolom. Di sini ditentukan secara tegas mana kolom teks yang bertindak sebagai fitur utama (prediktor) dan kolom kategorikal yang bertindak sebagai label/target (variabel dependen).

2.2. *Text Preprocessing* (Pemrosesan & Ekstraksi Fitur Teks)

Tahap ini krusial untuk mengubah data teks tak terstruktur menjadi data numerik terstruktur yang siap diproses algoritma. Prosesnya meliputi pembersihan teks (*case folding*, *tokenizing*, *stopwords removal*, dan *stemming/lemmatization*) melalui representasi *Corpus*. Selanjutnya, teks ditransformasikan menggunakan metode *Bag of Words* (atau TF-IDF) untuk mengekstrak bobot frekuensi kata menjadi matriks fitur.

2.3. *Modeling* (Pengambilan Sampel & Pelatihan Klasifikasi)

Data yang telah diekstrak fiturnya kemudian dibagi menggunakan teknik sampling (seperti pembagian data latih dan data uji). Setelah itu, penelitian menerapkan pendekatan

benchmarking dengan melatih lima algoritma klasifikasi *supervised learning* yang berbeda secara simultan, yaitu: *kNN*, *Naive Bayes*, *Decision Tree*, *Random Forest*, dan *Support Vector Machine (SVM)* untuk mempelajari pola pada data teks.

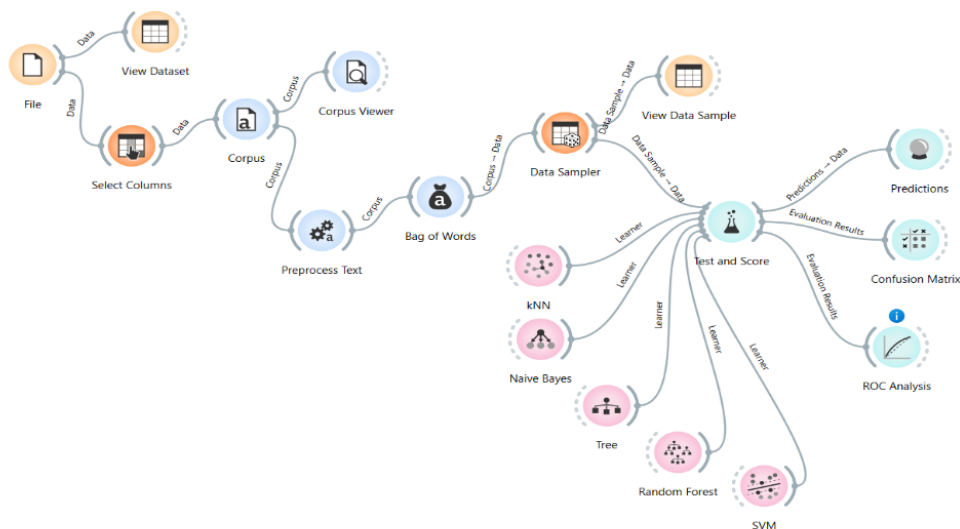
2.1. Evaluation (Pengujian & Validasi Performa)

Tahap akhir metode ini adalah menguji akurasi tebakan dan validitas dari kelima model yang telah dilatih. Menggunakan widget *Test and Score* (melalui metode seperti *Cross-Validation* atau *Random Sampling*), performa model diukur secara komparatif. Evaluasi diperdalam menggunakan instrumen statistik dan visual berupa tabel *Confusion Matrix*, *visualisasi Predictions*, dan grafik *ROC Analysis* untuk menentukan algoritma mana yang paling optimal dalam mengklasifikasikan data teks tersebut.

III. HASIL DAN PEMBAHASAN

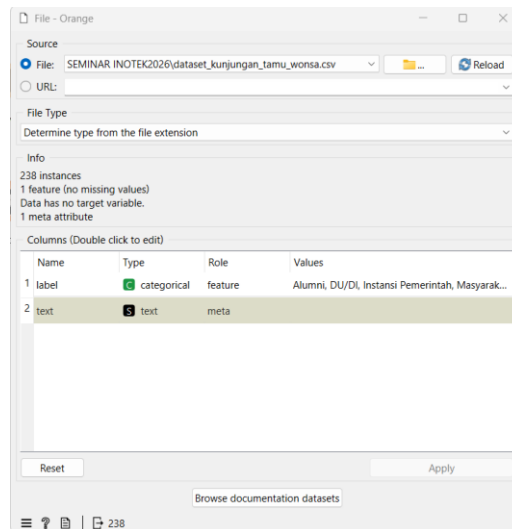
3.1. Skenario Penelitian

Dalam penelitian ini digunakan aplikasi Orange Data Mining versi 3.40. dimana desain penelitian adalah sebagai berikut:



Gambar 2. Desain Penelitian

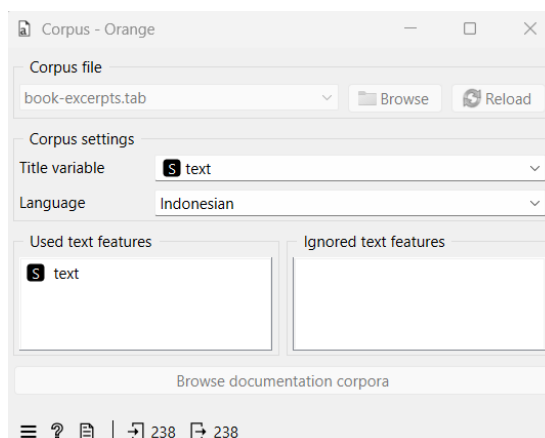
A. Pemasangan Dataset dan Pemilihan Kolom Fitur



Gambar 3. Tampilan *Widget File*

Jendela berikut menampilkan konfigurasi awal dari *widget File* di *Orange Data Mining* untuk memuat dataset eksternal bernama `dataset_kunjungan_tamu_wonsa.csv`. Dataset ini memiliki total 238 baris data (*instances*) tanpa ada data yang kosong (*no missing values*), serta terdiri dari satu kolom fitur dan satu atribut meta. Namun, pada bagian informasi sistem memberikan catatan bahwa data belum memiliki variabel target ("*Data has no target variable*"), yang berarti peran dari setiap kolom belum dikonfigurasi secara tepat untuk kebutuhan pemodelan prediktif atau klasifikasi terawasi (*supervised learning*).

B. *Corpus* (Data Kumpulan Kata)

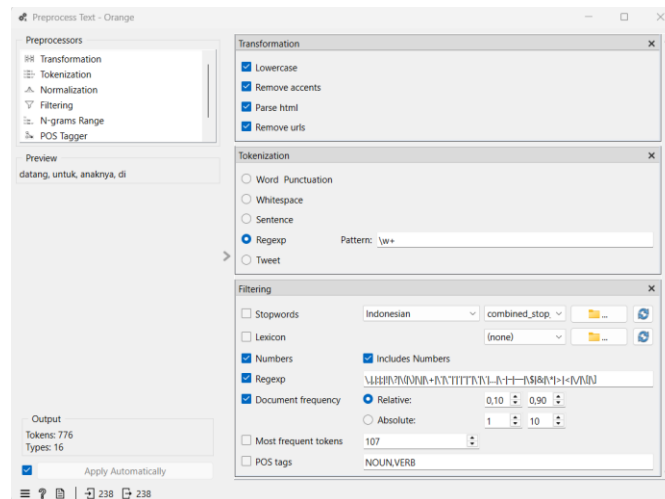


Gambar 4. *Corpus Data*

Jendela ini adalah konfigurasi *widget Corpus* di Aplikasi *Orange Data Mining* yang berfungsi untuk memuat dan mengatur dataset teks. File yang digunakan adalah `book-excerpts.tab`, yaitu dataset hasil dari pembuatan corpus dataset csv dalam format bawaan Orange. Pada bagian pengaturan, kolom "text" ditetapkan sebagai Title Variable sekaligus sebagai satu-satunya fitur teks yang aktif dianalisis, sementara bahasa dokumen

diset ke *Indonesian* agar proses pengolahan teks seperti tokenisasi dan penghapusan stopwords mengikuti aturan bahasa Indonesia. Tidak ada kolom teks yang diabaikan karena sisi *Ignored text features* masih kosong. Angka 238 di bagian bawah menunjukkan bahwa seluruh 238 dokumen berhasil dimuat dan diteruskan tanpa ada data yang gugur, artinya konfigurasi ini sudah berjalan dengan benar dan siap disambungkan ke widget berikutnya seperti *Preprocess Text*.

C. *Preprocessing Text* (Pengolahan Kata)



Gambar 5. *Preprocessing Data*

Jendela ini adalah konfigurasi widget *Preprocess Text* di Orange yang berfungsi untuk membersihkan dan mempersiapkan teks sebelum dianalisis lebih lanjut. Widget ini memiliki tiga tahap utama yang aktif, yaitu Transformation, Tokenization, dan Filtering.

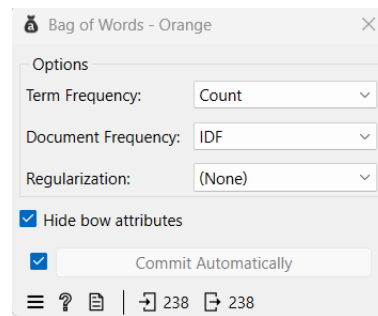
Pada tahap Transformation, semua opsi diaktifkan teks diubah menjadi huruf kecil seluruhnya (*Lowercase*), aksen atau karakter diakritik dihapus (*Remove accents*), tag HTML dibersihkan (*Parse HTML*), dan tautan URL dihilangkan (*Remove URLs*). Tujuannya adalah menyeragamkan bentuk teks agar lebih bersih sebelum diproses lebih lanjut.

Pada tahap Tokenization, metode yang dipilih adalah Regexp dengan pola `\w+`, artinya teks dipotong menjadi token berdasarkan urutan karakter alfanumerik saja. Tanda baca, simbol, dan spasi tidak dianggap sebagai token sehingga hanya kata-kata bermakna yang tersisa.

Pada tahap Filtering, stopwords bahasa Indonesia diset menggunakan daftar `combined_stop` namun centangannya tidak aktif, sehingga stopwords belum disaring, hal ini terlihat dari hasil preview di kiri yang masih memunculkan kata-kata umum seperti "untuk" dan "di". Filter angka diaktifkan (*Numbers + Includes Numbers*), filter karakter khusus melalui *Regexp* juga aktif untuk membuang simbol-simbol yang tidak relevan. *Document frequency* diaktifkan dengan rentang relatif 0,10 hingga 0,90, artinya token yang muncul di kurang dari 10% atau lebih dari 90% dokumen akan dibuang karena dianggap terlalu langka atau terlalu umum.

Hasil akhir yang terlihat di bagian Output menunjukkan 776 token dengan 16 tipe unik yang dihasilkan dari 238 dokumen yang masuk dan keluar. Angka tipe yang sangat kecil (16) kemungkinan karena filter document frequency yang cukup ketat memangkas banyak kata.

D. Bag of Word



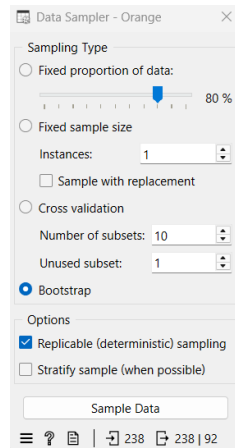
Gambar 6. *Bag of Words*

Jendela ini adalah konfigurasi *widget Bag of Words (BoW)* di *Orange Data Mining* yang berfungsi mengubah teks yang sudah diproses menjadi representasi numerik berupa matriks angka, sehingga data teks dapat digunakan oleh algoritma machine learning.

Pada pengaturan *Term Frequency* dipilih *Count*, artinya setiap kata dihitung berdasarkan frekuensi kemunculannya secara mentah dalam setiap dokumen. Jika kata "buku" muncul 3 kali dalam satu dokumen, maka nilainya adalah 3. Pada *Document Frequency* dipilih *IDF (Inverse Document Frequency)*, artinya bobot setiap kata disesuaikan berdasarkan seberapa jarang kata tersebut muncul di seluruh dokumen kata yang muncul di banyak dokumen mendapat bobot lebih rendah, sedangkan kata yang lebih jarang dan spesifik mendapat bobot lebih tinggi. Kombinasi *Count* dan *IDF* ini pada dasarnya menghasilkan skema pembobotan *TF-IDF*, yang umum digunakan dalam analisis teks karena mampu menonjolkan kata-kata yang benar-benar informatif. Sementara itu, *Regularization* diset ke *None*, artinya tidak ada normalisasi tambahan yang diterapkan pada vektor hasil BoW.

Opsi *Hide bow attributes* dicentang, yang berarti kolom-kolom hasil BoW disembunyikan dari tampilan tabel data agar tidak memenuhi layar, meskipun data tetap diteruskan secara penuh ke widget berikutnya. Angka 238 masuk dan 238 keluar di bagian bawah mengonfirmasi bahwa seluruh dokumen berhasil dikonversi menjadi representasi BoW tanpa ada yang gugur, dan widget siap meneruskan data numerik ini ke tahap berikutnya seperti Data Sampler atau langsung ke model klasifikasi.

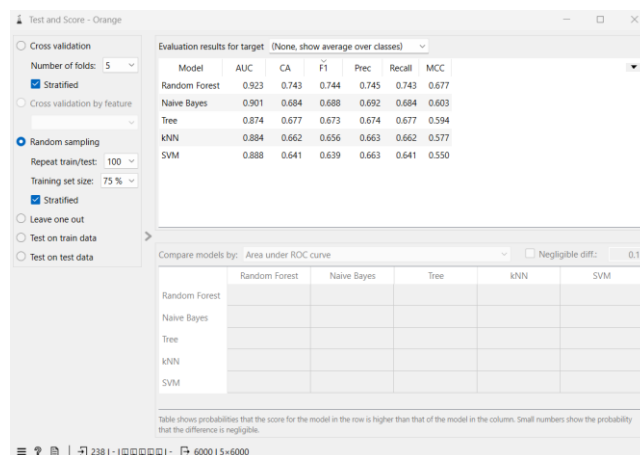
E. Sampling Data



Gambar 7. Data Sampling

Jendela ini adalah konfigurasi *widget Data Sampler* di Orange Data Mining yang berfungsi membagi dataset sebelum diuji oleh model. Metode yang dipilih adalah Bootstrap, yaitu teknik pengambilan sampel dengan cara mengambil data secara acak dengan pengembalian (*sampling with replacement*) dari dataset asli artinya satu data bisa terpilih lebih dari sekali sebagai data latih, sementara data yang tidak terpilih sama sekali digunakan sebagai data uji. Opsi *Replicable (deterministic)* sampling dicentang agar hasil pembagian data selalu sama setiap kali dijalankan, sehingga eksperimen dapat direproduksi. Angka di bagian bawah menunjukkan 238 dokumen masuk, dengan 238 dokumen sebagai data latih dan 92 dokumen sebagai data uji, yang merupakan hasil tipikal dari metode *Bootstrap* di mana sekitar 63% data unik terpilih sebagai *training set* dan sisanya menjadi *test set*.

F. Test and Scoring

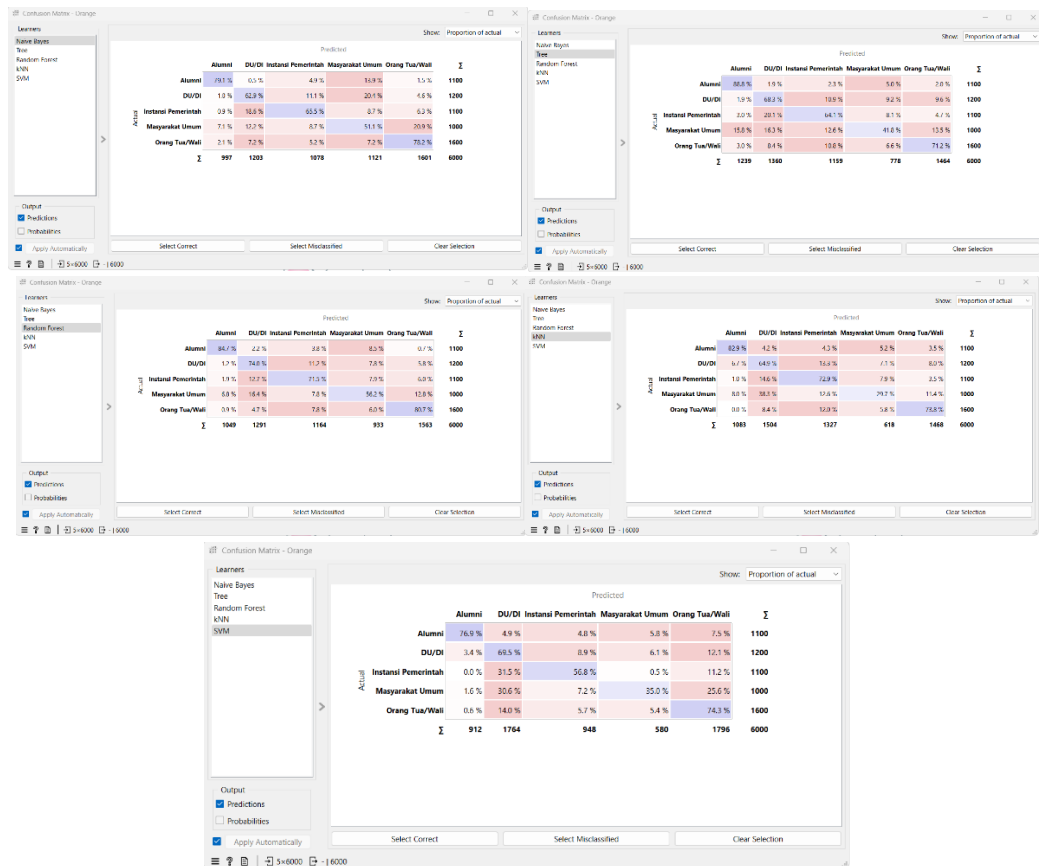


Gambar 8. Test & Score

Hasil evaluasi *Test and Score* menunjukkan bahwa *Random Forest* merupakan model terbaik dengan nilai AUC 0,923, *F1 Score* 0,744, *Precision* 0,745, *Recall* 0,743, dan *MCC* 0,677. Diikuti oleh *Naive Bayes* dengan AUC 0,901 dan *F1 Score* 0,688, kemudian *Decision Tree*, *kNN*, dan *SVM* di posisi berikutnya. Secara keseluruhan, semua model menghasilkan nilai AUC di atas 0,87 yang menandakan bahwa seluruh model mampu

mengklasifikasikan teks dengan cukup baik, namun *Random Forest* secara konsisten unggul di semua metrik evaluasi sehingga menjadi model paling optimal untuk dataset ini.

G. Confusion Matrix



Gambar 9. Confussion Matriks

Ringkasan hasil *Confusion Matrix* dari kelima model adalah sebagai berikut:

Random Forest menjadi model dengan performa terbaik secara keseluruhan, ditunjukkan dengan nilai akurasi diagonal yang tinggi dan konsisten di semua kelas Alumni 84,7%, DU/DI 74,0%, Instansi Pemerintah 71,5%, Masyarakat Umum 56,2%, dan Orang Tua/Wali 80,7%. Model ini paling stabil dalam mengklasifikasikan seluruh kategori dengan tingkat kesalahan yang relatif kecil.

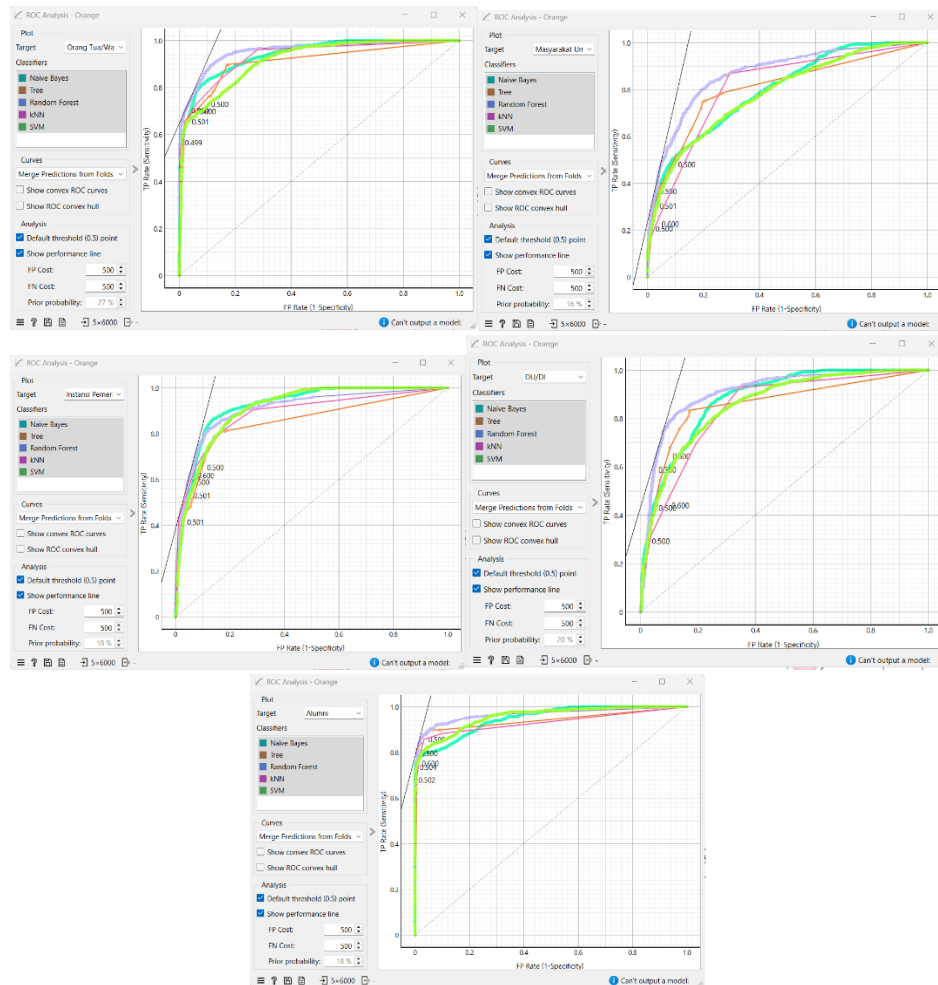
kNN menunjukkan performa baik pada kelas Alumni (82,9%), Instansi Pemerintah (72,9%), dan Orang Tua/Wali (73,8%), namun sangat lemah pada kelas Masyarakat Umum yang hanya mencapai 29,7% dan banyak salah diprediksi sebagai DU/DI.

Decission Tree unggul pada kelas Alumni (88,8%) namun memiliki kelemahan signifikan pada kelas Masyarakat Umum yang hanya 41,8%, menjadikannya model yang kurang seimbang antar kelas.

Naive Bayes memiliki akurasi yang cukup pada Alumni (79,1%) dan Orang Tua/Wali (78,2%), tetapi lemah pada Masyarakat Umum (51,1%) dengan banyak kesalahan prediksi ke kelas DU/DI dan Orang Tua/Wali.

SVM menjadi model dengan performa terburuk, terutama pada kelas Instansi Pemerintah (56,8%) dan Masyarakat Umum (35,0%) yang sangat rendah, serta banyak data yang salah diklasifikasikan ke kelas DU/DI.

H. ROC Analysis



Gambar 10. ROC Analysis

Hasil *ROC Analysis* menunjukkan bahwa *Random Forest* secara konsisten memiliki kurva ROC yang paling tinggi dan paling mendekati sudut kiri atas pada seluruh kelas target Alumni, DU/DI, Instansi Pemerintah, Masyarakat Umum, maupun Orang Tua/Wali yang mengonfirmasi bahwa model ini memiliki kemampuan diskriminasi terbaik dibanding model lainnya.

Naive Bayes juga menunjukkan performa yang kompetitif dengan kurva yang cukup tinggi di hampir semua kelas, meskipun sedikit di bawah *Random Forest*. Sementara *Decision Tree* dan *kNN* berada di posisi menengah dengan kurva yang bervariasi

tergantung kelasnya. *SVM* secara konsisten menampilkan kurva yang paling rendah dan paling jauh dari sudut kiri atas di semua kelas, yang sejalan dengan hasil *AUC* terendah pada evaluasi *Test and Score* sebelumnya.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma *Random Forest* terbukti menjadi algoritma terbaik dalam mengklasifikasikan kunjungan tamu pada institusi pendidikan dibandingkan dengan algoritma *Naive Bayes*, *Decision Tree*, *kNN*, dan *SVM*. Hal ini dibuktikan dengan nilai *AUC* sebesar 0,923, *F1 Score* sebesar 0,744, *Precision* sebesar 0,745, *Recall* sebesar 0,743, dan *MCC* sebesar 0,677 yang secara konsisten unggul di seluruh metrik evaluasi maupun pada setiap kelas target kunjungan tamu.

Penerapan algoritma *Random Forest* pada dataset kunjungan tamu institusi pendidikan menunjukkan bahwa model ini mampu mengenali pola teks dari berbagai kategori pengunjung dengan tingkat akurasi yang baik, sehingga berpotensi diimplementasikan sebagai sistem klasifikasi otomatis kunjungan tamu di sekolah maupun institusi pendidikan lainnya. Kemampuan ini dapat memberikan kontribusi nyata dalam efisiensi pengelolaan administrasi tamu, khususnya dalam mempercepat proses identifikasi dan pencatatan jenis kunjungan secara otomatis berbasis teks.

Sebagai saran, penelitian selanjutnya dapat mempertimbangkan penambahan jumlah data dan variasi dokumen teks agar model yang dihasilkan lebih robust, terutama untuk kelas Masyarakat Umum yang pada penelitian ini masih menunjukkan tingkat klasifikasi paling rendah di antara kelas lainnya. Selain itu, eksplorasi terhadap teknik *preprocessing* yang lebih lanjut seperti pengaktifan *stopword removal* bahasa Indonesia serta penggunaan metode ekstraksi fitur yang lebih canggih dapat menjadi langkah pengembangan untuk meningkatkan performa klasifikasi secara keseluruhan.

DAFTAR PUSTAKA

- [1] A. M. Majid and I. Nawangsih, "Perbandingan Metode Ensemble Untuk Meningkatkan Akurasi Algoritma Machine Learning Dalam Memprediksi Penyakit Breast Cancer (Kanker Payudara)," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 23, no. 1, pp. 97–104, Feb. 2024, doi: 10.53513/JIS.V23I1.9563.
- [2] S. Adnin Kamila, R. Sri Sulistijowati, I. Susanto, and F. Matematika dan Ilmu Pengetahuan Alam, "Classification of Heart Disease Using Decision Tree and Random Forest," *Semin. Nas. Teknol. Sains*, vol. 2, no. 1, pp. 7–12, Jan. 2023, doi: 10.29407/STAINS.V2I1.2816.
- [3] P. B. Alamin and) Diah Aryani, "PERBANDINGAN ALGORITMA KNN, NAIVE BAYES DAN SVM UNTUK KLASIFIKASI STATUS STUNTING ANAK BALITA DI PUSKESMAS SEPATAN," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 4, no. 4, pp. 360–369, Nov. 2025, doi: 10.55123/STORAGE.V4I4.6591.
- [4] H. T. Santoso, F. A. Felmidi, A. N. Fadhila, A. Ristyawan, and E. Daniati, "Analisis Kinerja Algoritma Data Mining pada Klasifikasi Tingkat Obesitas dengan K-Fold Cross Validation dan AUC," *Pros. SEMNAS INOTEK (Seminar Nas. Inov. Teknol.)*, vol. 8, no. 1, pp. 113–122, Jul. 2024, doi: 10.29407/INOTEK.V8I1.4917.
- [5] A. Perbandingan Algoritma Classification Berdasarkan Komposisi Label, T. Krisna Amarya, A. G. Candra Andy, R. Achmad, E. Daniati, and A. Ristyawan, "Analisa Perbandingan Algoritma Classification Berdasarkan Komposisi Label," *Pros. SEMNAS INOTEK (Seminar Nas. Inov. Teknol.)*, vol. 8, no. 1, pp. 32–40, Jul. 2024, doi: 10.29407/INOTEK.V8I1.4906.
- [6] D. Safitri, S. S. Hilabi, and F. Nurapriani, "ANALISIS PENGGUNAAN ALGORITMA KLASIFIKASI DALAM PREDIKSI KELULUSAN MENGGUNAKAN ORANGE DATA MINING," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 8, no. 1, pp. 75–81, Jan. 2023, doi: 10.36341/RABIT.V8I1.3009.
- [7] Z. Amri, K. Kusriani, and K. Kusnawi, "Prediksi Tingkat Kelulusan Mahasiswa menggunakan Algoritma Naïve Bayes, Decision Tree, ANN, KNN, dan SVM," *Edumatic J. Pendidik. Inform.*, vol. 7, no. 2, pp. 187–196, Dec. 2023, doi: 10.29408/edumatic.v7i2.18620.
- [8] I. Setiawan, I. Fatah Yasin, Y. Tri Desianti, P. Studi Sistem Dan Teknologi Informasi, F. Sains Dan Teknologi, and A. Surakarta, "Komparasi Kinerja Algoritma Random Forest, Decision Tree, Naïve Bayes, dan KNN dalam Prediksi Tingkat Depresi Mahasiswa menggunakan Student Depression Dataset," *J. Ilmu Komput. dan Teknol.*, vol. 6, no. 1, pp. 47–58, Feb. 2025, doi: 10.35960/IKOMTI.V6I1.1756.
- [9] P. Sejati, M. Munawar, M. Pilliang, and H. Akbar, "Studi Komparasi Naive Bayes, K-Nearest Neighbor, dan Random Forest untuk Prediksi Calon Mahasiswa yang Diterima atau Mundur," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 7, pp. 1341–1348, Dec. 2022, doi: 10.25126/JTIK.2022976737.
- [10] A. A. Karim, M. A. Prasetyo, and M. R. Saputro, "Perbandingan Metode Random Forest, K-Nearest Neighbor, dan SVM Dalam Prediksi Akurasi Pertandingan Liga Italia," *Semin. Nas. Teknol. Sains*, vol. 2, no. 1, pp. 377–382, Jan. 2023, doi: 10.29407/STAINS.V2I1.2877.
- [11] A. Laksono and F. A. Rahmadani, "Analisis Sentimen Opini Publik di platform X (Twitter) terhadap Bencana Banjir Sumatera," *Semin. Nas. Teknol. Sains*, vol. 5, no. 1, pp. 689–698, Jan. 2026, doi: 10.29407/XQ0RVZ75.